Anomaly Detection with the Switching Kalman Filter for Structural Health Monitoring

Luong Ha Nguyen* and James-A. Goulet
Department of Civil, Geologic and Mining Engineering
Ecole Polytechnique de Montreal, CANADA

November 30, 2017

Abstract

Detecting changes in structural behaviour, i.e. anomalies over time is an important aspect in structural safety analysis. The amount of data collected from civil structures keeps expanding over years while there is a lack of data-interpretation methodology capable of reliably detecting anomalies without being adversely affected by false alarms. This paper proposes an anomaly detection method that combines the existing Bayesian Dynamic Linear Models framework with the Switching Kalman Filter theory. The potential of the new method is illustrated on the displacement data recorded on a dam in Canada. The results show that the approach succeeded in capturing the anomalies caused by refection work without triggering any false alarms. It also provided the specific information about the dam's health and conditions. This anomaly detection method offers an effective data-analysis tool for Structural Health Monitoring.

Keywords: Anomaly Detection, Bayesian, Dynamic Linear Models, Switch Kalman Filter, Structural Health Monitoring, False Alarm, Dam.

1 Introduction

Around the world, civil structures are in poor condition [1,2]. As early as in the 1970s many structures have been monitored to improve the understanding of their behaviour [3,4]. This research field is known as *Structural Health Monitoring* (SHM). Quantities monitored on a structure are typically displacements, strains, inclination or accelerations [5,6]. Sensing technology has evolved over the last decades and is now cheap and widely available. The hardware developments led to an increase in the amount of available data. In this paper, we focus on long-term condition SHM. Many methodologies from the field of applied statistics and machine learning [7,8] have been proposed for interpreting the time-series data in order to deduce valuable insights based on these data.

A first key aspect limiting the applicability of SHM is that there is currently a lack of data-interpretation methodology capable of reliably detecting anomalies in time series without also being adversely affected by *false alarms*. An *anomaly* is defined here as a change

*Corresponding author: luongha.nguyen@gmail.com

in the behaviour of a structure. A second key aspect is that in order to be financially viable for practical applications, data-interpretation methods must be easily transferable from one structure to another and from one measurement type to another. This second aspect is mandatory if the objective is to deploy SHM systems across populations of structures. A third key aspect is that no training sets with labeled conditions (normal and abnormal) are available.

Existing *Regression Methods* (RMs) for SHM model the dependence between observed structural responses and time-dependent covariates such as temperature and loading. In the field of dam engineering, the most common regression method is the HST (Hydrostatic, Seasonal, Time) method [9–11] and analogue derivations [12–14] that have been applied to interpret dam behaviour through displacement, pressure, and flow-rate data. In addition to the HST, Neural Network [15], Support Vector Machines [16, 17], Boosted Regression Trees [18] and others [19, 20] are used for the same purpose. The first drawback of these RMs is that once the model is built using a training set, it stops evolving as new data is collected. The second drawback is that anomaly detection is based on a *hypothesis-testing* procedure. A probability density function of the error between observation and the model prediction is identified for the training set and then employed to detect anomalies. The presence of the anomalies is tested based on the distance between the training set and test-set confidence regions. This procedure tends to be prone to false alarms in the presence of outliers.

Contrary to common RMs, *State Space Models* (SSMs) continue learning from the new data after the training set. In the SSMs, the structural responses are modelled by the superposition of hidden states that are not directly observed. An example for the SSMs is Autoregressive Models (ARs) that are employed to classify damage scenarios on experimental data for the IASC–ASCE benchmark a four-storey frame structure, the Z24 bridge in Switzerland and the Malaysia–Singapore Second Link bridge [21]. The ARs are also applied to damage detection on the Steel-Quake structure at the Joint Research Center in Ispra (Italy) [22]. The limited predictive capacity of ARs hinders their widespread applicability. Other dynamic modelling methods based on Kalman Filter variants [23, 24] are used to identify changes in the modal parameters such as the stiffness and damping for detecting anomalies. Such as typically require detailed information about a structure, which is not suited for a widespread deployment across thousands of bridges and dams that are all different from one to another.

Bayesian Dynamic Linear Models (BDLMs) [25] based on the SSMs have shown to be a promising solution in order to address the above limitations. The idea behind BDLM is that the observed structural observation is decomposed into a set of hidden components. The generic hidden components can be among others, a *Local Level* to describe the baseline response of structures, a *Periodic Component* to describe periodic effects such as temperature, an *Autoregressive Component* to capture time-dependent model approximation errors. In the BDLM, the rate of changes in the evolution of the Local Level is characterized by a *Local Trend*. If any changes occur in the Local Trend, a *Local Acceleration* component must be added to model its rate of change. The key challenge here is that in its current form, the BDLM can only model behaviour of structures under *stationary*, i.e. *normal* conditions. In order to detect the occurrence of anomalies, it needs to be extended to operate in *non-stationary*, i.e. *abnormal*, conditions.

This paper proposes an anomaly detection method that combines the existing BDLM with the *Switching Kalman Filter* (SKF) theory [26]. In the field of machine learning, the SKF is used in many case studies [27–29] for handling non-stationary conditions. The key features of the approach proposed is that:

- It enables early anomaly detection
- It is robust towards false alarms in real operation condition
- It does not require labeled training data with normal and abnormal conditions.

The paper is organized as follows. The Section 2 presents a summary of the SKF theory. The Section 3 describes the methodology for anomaly detection. The section 4 illustrates the potential of the new approach on displacement data recorded on a dam located in Canada.

2 Switching Kalman Filter

This section presents the mathematical formulations for the combination of existing *Bayesian Dynamic Linear Model* with the *Switching Kalman Filter*. A BDLM is defined by the following linear equations:

Observation equation

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t, \quad \begin{cases} \mathbf{y}_t \sim \mathcal{N}(\mathbb{E}[\mathbf{y}_t], \mathrm{cov}[\mathbf{y}_t]) \\ \mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_{\mathbf{t}}) \\ \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathbf{t}}) \end{cases} \tag{1}$$

Transition equation

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \{\ \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\mathbf{t}}), \tag{2}$$

where $\mathbf{y}_t$ is the observation vector at the time $t \in (1:T)$, $\mathbf{C}_t$ is the observation matrix , $\mathbf{x}_t$ is the hidden state variables that they are not directly observed, $\mathbf{v}_t$ is the Gaussian measurement error with mean zero and covariance matrix $\mathbf{R}_t$, $\mathbf{A}_t$ is the transition matrix, and $\mathbf{w}_t$ is the Gaussian model error with mean zero and covariance matrix $\mathbf{Q}_t$. Equations 1 and 2 estimated using the Kalman Filter [30]. The specificity of BDLM is to build model matrices $\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t$ using a pre-defined sub-component structure [25]. The SKF enables to model the different states of a system, each having its own set of model matrix by estimating, over time steps, the probability of multiple model classes. Following the notation from Murphy [26], the SKF algorithm is divided into the *filter* and *collapse* steps.

SKF-Filter step

The SKF-Filter step is equivalent to the *Kalman filter* employed for the existing BDLM. However, the notation for Kalman filter (KF) algorithm needs to be adapted to include the *Markov-switching variable* $s_t \in \{1, 2, \ldots, \mathtt{S}\}$, each one corresponding to a distinct *filtering*

model defined by its model matrices. The Markov-switching variables at time t and $t-1$ are respectively $s_{t-1}=i$ and $s_t=j$. The *superscript* inside the parentheses $^{i(j)}$ is employed to denote the current state j at the time t given the state i at time $t-1$. For the SKF, the prediction and measurement steps from the Kalman filter (KF) algorithm are rewritten as

KF-Prediction step

$$\begin{aligned}
p\left(\mathbf{x}_t^{i(j)}|\mathbf{y}_{1:t-1}\right) &= \mathcal{N}\left(\mathbf{x}_t^{i(j)};\boldsymbol{\mu}_{t|t-1}^{i(j)},\boldsymbol{\Sigma}_{t|t-1}^{i(j)}\right) && \text{Prior state estimate}\\
\boldsymbol{\mu}_{t|t-1}^{i(j)} &\triangleq \mathbf{A}_t^{i(j)}\boldsymbol{\mu}_{t-1|t-1}^{i} && \text{Prior expected value}\\
\boldsymbol{\Sigma}_{t|t-1}^{i(j)} &\triangleq \mathbf{A}_t^{i(j)}\boldsymbol{\Sigma}_{t-1|t-1}^{i}\left(\mathbf{A}_t^{i(j)}\right)^{\mathsf{T}}+\mathbf{Q}_t^{i(j)} && \text{Prior covariance}
\end{aligned}$$

KF-Measurement step

$$\begin{aligned}
p\left(\mathbf{x}_t^{i(j)}|\mathbf{y}_{1:t}\right) &= \mathcal{N}\left(\mathbf{x}_{t|t}^{i(j)};\boldsymbol{\mu}_{t|t}^{i(j)},\boldsymbol{\Sigma}_{t|t}^{i(j)}\right) && \text{Posterior state estimate}\\
\boldsymbol{\mu}_{t|t}^{i(j)} &= \boldsymbol{\mu}_{t|t-1}^{i(j)}+\mathbf{K}_t^{i(j)}\mathbf{r}_t^{i(j)} && \text{Posterior expected value}\\
\boldsymbol{\Sigma}_{t|t}^{i(j)} &= \left(\mathbf{I}-\mathbf{K}_t^{i(j)}\mathbf{C}_t^{i(j)}\right)\boldsymbol{\Sigma}_{t|t-1}^{i(j)} && \text{Posterior covariance}\\
\mathbf{r}_t^{i(j)} &\triangleq \mathbf{y}_t-\hat{\mathbf{y}}_t^{i(j)} && \text{Innovation vector}\\
\hat{\mathbf{y}}_t^{i(j)} &\triangleq \mathbb{E}[\mathbf{y}_t|\mathbf{y}_{1:t-1}]=\mathbf{C}_t^{i(j)}\boldsymbol{\mu}_{t|t-1}^{i(j)} && \text{Predicted observations vector}\\
\mathbf{K}_t^{i(j)} &\triangleq \boldsymbol{\Sigma}_{t|t-1}^{i(j)}\left(\mathbf{C}_t^{i(j)}\right)^{\mathsf{T}}\left(\mathbf{G}_t^{i(j)}\right)^{-1} && \text{Kalman gain matrix}\\
\mathbf{G}_t^{i(j)} &\triangleq \mathbf{C}_t^{i(j)}\boldsymbol{\Sigma}_{t|t-1}^{i(j)}\left(\mathbf{C}_t^{i(j)}\right)^{\mathsf{T}}+\mathbf{R}_t^{i(j)} && \text{Innovation covariance matrix.}
\end{aligned}$$

The *Kalman gain* matrix $\mathbf{K}_t^{i(j)}$ represents the relative importance of the innovation vector $\mathbf{r}_t^{i(j)}$ with regard to the prior expected value $\boldsymbol{\mu}_{t|t-1}^{i,(j)}\equiv\mathbb{E}[\mathbf{x}_t^{i(j)}|\mathbf{y}_{1:t-1}]$. The model uncertainty is described by the model error covariance matrix $\mathbf{Q}_t^{i(j)}$ which depends on the state i at time $t-1$ and the state j at time t. For the case where there is no state transition between $t-1$ and t, model classes are assumed to be dependent upon the arrival state j at time t so that $\mathbf{Q}_t^{i(j)}=\mathbf{Q}_t^{j}$. If between time steps $t-1$ and t there is a transition from one state to another, the matrix $\mathbf{Q}_t^{i(j)}$ needs to be identified i.e $\mathbf{Q}_t^{i(j)}\neq\mathbf{Q}_t^{j}$. For common cases, matrices defining the transition and observation models are only dependent on the arrival state j at time t,

$$\mathbf{A}_t^{i(j)}=\mathbf{A}_t^{j},\quad \mathbf{C}_t^{i(j)}=\mathbf{C}_t^{j},\quad \mathbf{R}_t^{i(j)}=\mathbf{R}_t^{j}.$$

The Kalman filter algorithm described above is summarized in its short form as

$$(\boldsymbol{\mu}_{t|t}^{i(j)},\boldsymbol{\Sigma}_{t|t}^{i(j)},\mathcal{L}_t^{i(j)}) = \text{Filter}(\boldsymbol{\mu}_{t-1|t-1}^{j},\boldsymbol{\Sigma}_{t-1|t-1}^{i},\mathbf{A}_t^{j},\mathbf{C}_t^{j},\mathbf{Q}_t^{i(j)},\mathbf{R}_t^{j}) \tag{3}$$

where $\mathcal{L}_t^{i(j)}$ measures the likelihood that the state at time $t-1$ was $s_{t-1}=i$ and that it switches to $s_t=j$. The likelihood of such as switch $\mathcal{L}_t^{i(j)}$ is defined as

$$\begin{aligned}\mathcal{L}_t^{i(j)} &= p(\mathbf{y}_t|s_t=j, s_{t-1}=i, \mathbf{y}_{1:t-1}) \\ &= \mathcal{N}(\mathbf{y}_t; \mathbf{C}_t^j \boldsymbol{\mu}_{t|t-1}^{i(j)}, \mathbf{R}_t^j + \boldsymbol{\Sigma}_{t|t-1}^{i(j)}).\end{aligned}$$

Note that the Filter step presented in Equation 3 can either be performed using the Kalman method as presented above or using the UD filter [25]. The UD method is equivalent to the Kalman method, yet it is numerically more stable [31].

SKF-Collapse step

The mean vector $\boldsymbol{\mu}_{t|t}^j \equiv \mathbb{E}[\mathbf{x}_t^j|\mathbf{y}_{1:t}]$ and covariance matrix $\boldsymbol{\Sigma}_{t|t}^j \equiv \text{cov}[\mathbf{x}_t^j|\mathbf{y}_{1:t}]$ are computed by collapsing the outputs from the filtering models according to their previous state probability, likelihood and transition probabilities. In order to further describe the collapse step, we introduce the notation

$$\begin{aligned}\Pr(s_{t-1}=i|\mathbf{y}_{1:t-1}) &= \pi_{t-1|t-1}^i && \text{Previous state probability} \\ \Pr(s_t=j|s_{t-1}=i) &= Z^{i(j)} && \text{Transition probability} \\ \Pr(s_{t-1}=i, s_t=j|\mathbf{y}_{1:t}) &= M_{t-1,t|t}^{i(j)} && \text{Joint probability} \\ \Pr(s_{t-1}=i|s_t=j, \mathbf{y}_{1:t}) &= W_{t-1|t}^{i(j)} && \text{State switching probability.}\end{aligned}$$

The joint probability of $s_t=j$ and $s_{t-1}=i$, given $\mathbf{y}_{1:t}$ is evaluated as

$$M_{t-1,t|t}^{i(j)} = \frac{\mathcal{L}_{t|t}^{i(j)} \cdot Z^{i(j)} \cdot \pi_{t-1|t-1}^i}{\sum_i \sum_j \mathcal{L}_{t|t}^{i(j)} \cdot Z^{i(j)} \cdot \pi_{t-1|t-1}^i}, \tag{4}$$

The denominator of Equation 4 is a normalization constant ensuring that $\sum_i \sum_j M_{t-1,t|t}^{i(j)} = 1$. The marginal probability of $s_t=j$ is obtained through marginalization following

$$\pi_{t|t}^j = \sum_i M_{t-1,t|t}^{i(j)}. \tag{5}$$

The collapsed mean vector $\boldsymbol{\mu}_{t|t}^j$ and covariance matrix $\boldsymbol{\Sigma}_{t|t}^j$ are defined as a *Gaussian mixture* so that

$$\begin{aligned}W_{t-1|t}^{i(j)} &= \frac{M_{t-1,t|t}^{i(j)}}{\pi_{t|t}^j} \\ \boldsymbol{\mu}_{t|t}^j &= \sum_i \boldsymbol{\mu}_{t|t}^{i(j)} \cdot W_{t-1|t}^{i(j)} \\ \mathbf{m} &= \boldsymbol{\mu}_{t|t}^{i(j)} - \boldsymbol{\mu}_{t|t}^j \\ \boldsymbol{\Sigma}_{t|t}^j &= \sum_i \left[W_{t-1|t}^{i(j)} \cdot (\boldsymbol{\Sigma}_{t|t}^{i(j)} + \mathbf{m}\mathbf{m}^\intercal) \right].\end{aligned} \tag{6}$$

The short-form notation for the collapse step is

$$(\boldsymbol{\mu}^{j}_{t|t}, \boldsymbol{\Sigma}^{j}_{t|t}, \pi^{j}_{t|t}) \quad = \quad \text{Collapse}(\boldsymbol{\mu}^{i(j)}_{t|t}, \boldsymbol{\Sigma}^{i(j)}_{t|t}, \mathbf{W}^{i(j)}_{t-1|t}).$$

An illustration of the SKF-filer and -collapse steps employed for describing the transition between two possible models is presented in Figure 1. The goal is to evaluate the mean vector

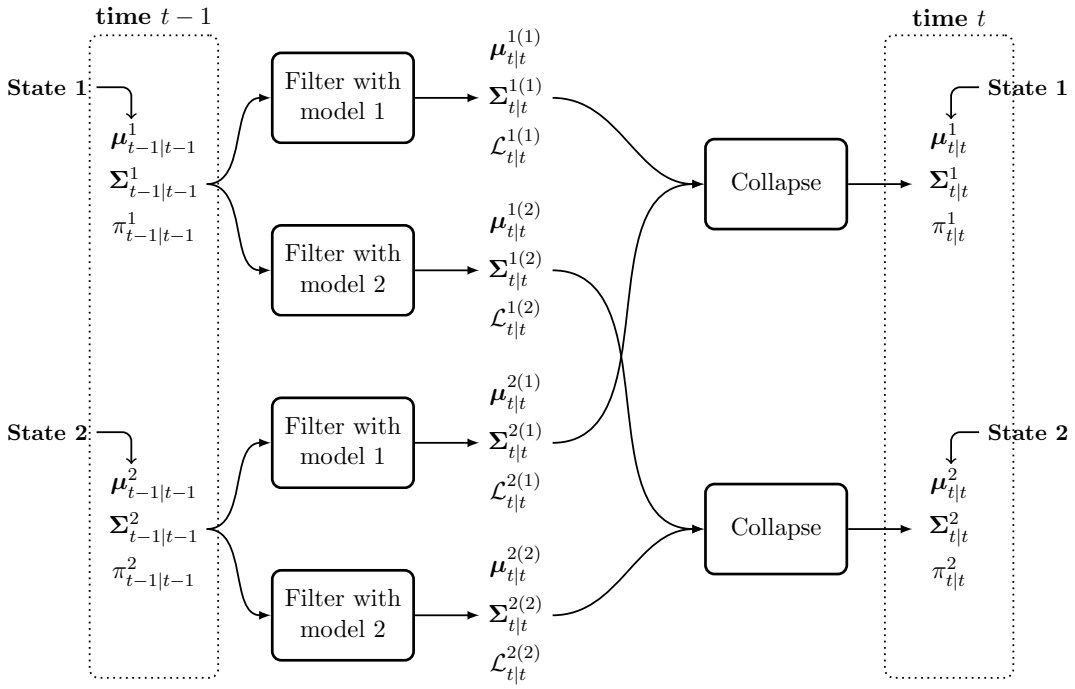

Figure 1: Illustration of the SKF algorithm for two states each having its own transition model. $^{(\cdot)}$ indicates the filtering model being used for computation.

$\boldsymbol{\mu}^{j}_{t|t}$ and the covariance matrix $\boldsymbol{\Sigma}^{j}_{t|t}$ for each model, $j \in \{1, 2\}$ along with the probability $\pi^{j}_{t|t}$ of each model at the time t, given the same two models at time $t-1$. When going from $t-1$ to t, there are four possibilities of transitions from a starting state $s_{t-1} = i$ to an arrival state $s_t = j$, each leading to its own mean vector $\boldsymbol{\mu}^{i(j)}_{t|t}$, covariance matrix $\boldsymbol{\Sigma}^{i(j)}_{t|t}$ and likelihood $\mathcal{L}^{i(j)}_{t|t}$. In the collapse step, the prior probability of each origin state is combined with the transition probability and the likelihood of each transition using Equation 4. The end result of the collapse step is a mean vector, covariance matrix and a probability for each model.

Combining the BDLM framework with SKF involves several unknown parameters that need to be learned from data. In common SHM applications, only structural responses $\mathbf{y}_{1:t}$ are available and the state of the structure s_t remains a hidden variable (i.e. non-observed). Therefore, the task of inferring s_t from $\mathbf{y}_{1:t}$ can be categorized as *semi-supervised* learning [32]. In that context, the set of parameters $\mathcal{P}^*$ is estimated employing the *Maximum*

Likelihood Estimation (MLE) technique where the log-likelihood function is

$$\begin{aligned}
\ln(\mathbf{y}_{1:T}|\mathcal{P}) &= \sum_{t=1}^{T} \ln p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\mathcal{P}) \\
&= \sum_{t=1}^{T} \ln \left[\sum_{j=1}^{\mathtt{s}} \sum_{i=1}^{\mathtt{s}} p(\mathbf{y}_t, s_t=j, s_{t-1}=i|\mathbf{y}_{t-1},\mathcal{P}) \right] \\
&= \sum_{t=1}^{T} \ln \left[\sum_{j=1}^{\mathtt{s}} \sum_{i=1}^{\mathtt{s}} p(\mathbf{y}_t|s_t=j, s_{t-1}=i, \mathbf{y}_{t-1},\mathcal{P}) \times p(s_t=j, s_{t-1}=i|\mathbf{y}_{t-1},\mathcal{P}) \right] \\
&= \sum_{t=1}^{T} \ln \left[\sum_{j=1}^{\mathtt{s}} \sum_{i=1}^{\mathtt{s}} p(\mathbf{y}_t|s_t=j, s_{t-1}=i, \mathbf{y}_{t-1},\mathcal{P}) \times p(s_t=j|s_{t-1}=i) \dots \right. \\
&\qquad \left. \dots \times p(s_{t-1}=i|\mathbf{y}_{1:t-1},\mathcal{P}) \right] \\
&= \sum_{t=1}^{T} \ln \left[\sum_{j=1}^{\mathtt{s}} \sum_{i=1}^{\mathtt{s}} \mathcal{L}_t^{i(j)} \cdot Z^{i(j)} \cdot \pi_{t-1|t-1}^{i} \right].
\end{aligned} \tag{7}$$

In this paper, the *Newton-Raphson* algorithm [33] is employed for estimating

$$\mathcal{P}^* = \arg\max_{\mathcal{P}} \ln(\mathbf{y}_{1:T}|\mathcal{P}).$$

3 Methodology for Anomaly Detection

This section presents an *Offline Batch Procedure* (OffBP) for anomaly detection applicable for full-scale structures such as bridges and dams. The OffBP employs the entire dataset for estimating the model parameters $\mathcal{P}$ using the MLE presented in Section 2. Because class labels are not observed, only the structural responses are employed in a semi-supervised learning context. In the OffBP, model parameters are assumed to be constant over time.

The key part of the OffBP anomaly detection lie in the model architecture employed for each state as well as in the transition probability matrix describing the switch between states. For the model architecture, each state has its own transition matrix $\mathbf{A}_t$ and model error covariance matrix $\mathbf{Q}_t$. According to the SKF theory in Section 2, the parameters from the matrix $\mathbf{Q}_t^{i(j)}$ need to be identified both for the stationary cases, i.e. $i=j$, and in the case of a state transition, i.e. $i \neq j$. In common case, a state transition is defined by a change in velocity and acceleration in the baseline behaviour. Therefore, the state transition is considered to only affect the *baseline behaviour* of the structure, which is separated from the external effects such as temperature and loading. The baseline behaviour regroups the local level, trend, and acceleration components. In the presence of a state transition, the model architecture must allow for an increase in the uncertainty for the local trend and acceleration components. For this purpose, the standard deviations in $\mathbf{Q}_t^{i(j),\,\text{baseline}}$ are

treated as an unknown parameter to be inferred from observations. An example for such as this case will be illustrated in Section 4.2. The transition probability matrix $\mathbf{Z}_t$ is identified based on the number of states. For $\mathtt{S}$ states $s_t \in \{=1, 2, 3 \cdots, \mathtt{S}\}$, the matrix $\mathbf{Z}_t$ is defined as

$$\mathbf{Z}_t = \begin{bmatrix} Z^{11} & Z^{12} & \dots & Z^{1\mathtt{S}} \\ Z^{21} & Z^{22} & \dots & Z^{2\mathtt{S}} \\ \vdots & \vdots & \ddots & \vdots \\ Z^{\mathtt{S}1} & Z^{\mathtt{S}2} & \dots & Z^{\mathtt{SS}} \end{bmatrix}$$

where $Z^{ij} = Pr(s_t = j | s_{t-1} = i)$ with $\sum_{j=1}^{\mathtt{S}} Z^{ij} = 1$.

The performance of the Newton-Raphson (NR) algorithm in Section 2 for learning parameters depends on (1) the initial parameter values $\mathcal{P}^0$ and (2) the initial mean $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{\Sigma}_0$ for the hidden states. The log-likelihood function is usually non-convex; poor guesses for either initial parameter values or hidden state initial values are prone to lead to a local maximum. In the case of anomaly detection, such a local maximum can trigger false alarms. In order to overcome this limitation, random sets of initial parameter values should be tested during the optimization procedure to ensure proper initial values. The second limitation is addressed using what we define as the *multi-pass* technique. The multi-pass recursively employs the Switch Kalman Smoother (SKS) [26] for estimating $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{\Sigma}_0$ for hidden states. Figure 2 illustrates the utilization of the multi-pass in the OffBP. During training, the model is first built using the initial parameter values $\mathcal{P}^n$,

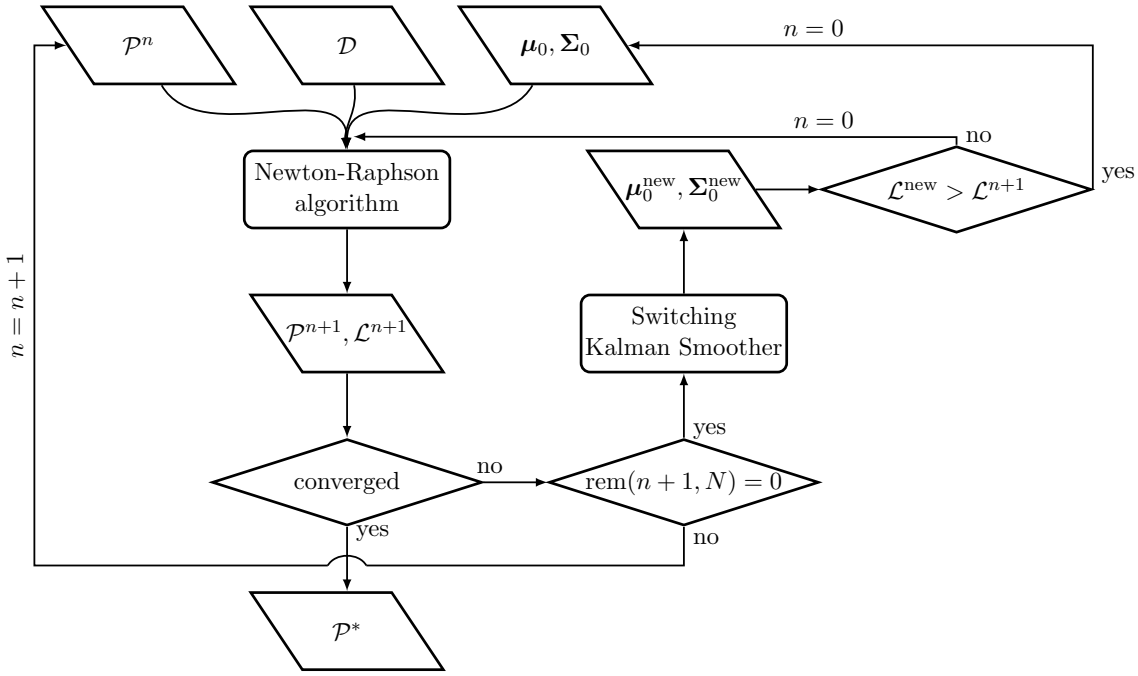

Figure 2: Illustration of the offline batch procedure using the multi-pass. rem$(n + 1, N)$ indicates the remainder after dividing N into $n + 1$.

initial values for hidden states $\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\}$, and the training data $\mathcal{D}$. The NR algorithm is thus employed for estimating the model parameters $\mathcal{P}^{n+1}$ and the corresponding log-likelihood $\mathcal{L}^{n+1}$. After N iterations, $\boldsymbol{\mu}_0^{\text{new}}$ and $\boldsymbol{\Sigma}_0^{\text{new}}$ are estimated using SKS and the index n is reset to 0. To be accepted as new initial values, the log-likelihood evaluated using $\{\boldsymbol{\mu}_0^{\text{new}}, \boldsymbol{\Sigma}_0^{\text{new}}\}$ needs to be greater than $\mathcal{L}^{n+1}$. This procedure is repeated until the convergence criterion is reached. The final output of the procedure is the set of optimal parameters $\mathcal{P}^*$. For the practical applications, N should be chosen so that the convergence criteria is not met before reaching N iterations. In common cases, N increases with the number of parameters to be estimated. In order to increase efficiency, the amount of data employed for estimating the initial values $\boldsymbol{\mu}_0^{\text{new}}$ and $\boldsymbol{\Sigma}_0^{\text{new}}$ can be smaller than the training data employed for the parameter optimization procedure.

4 Case-Study

In this study, the approach proposed for anomaly detection is applied to the displacement data collected on a dam located in Canada. The sensor studied is located on the west bank of the dam as shown in Figure 3. The displacement of the dam is monitored by an

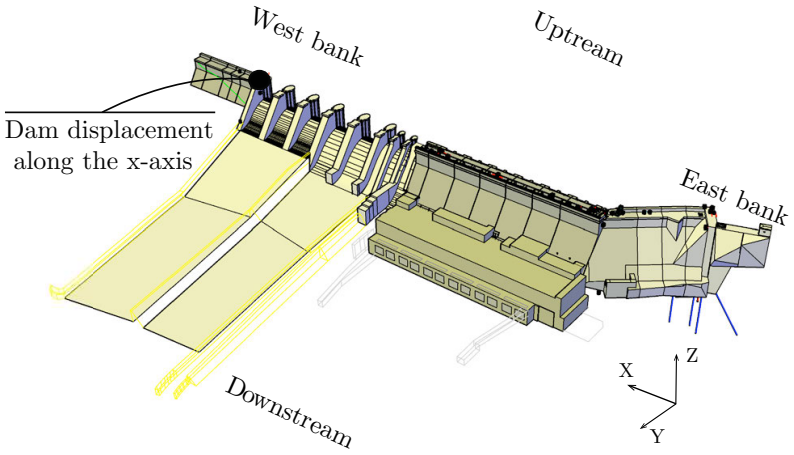

Figure 3: Location plan of sensors deployed across the structure to monitor the dam behaviour.

inverted pendulum system that provides the measurements in three orthogonal directions. The X-direction points toward the West Bank, the Y-direction follows the water flow and the Z-direction points upward.

4.1 Data Description

In order to examine the potential of the proposed anomaly detection method, this paper studies the horizontal displacement data along the X-direction recorded over the period of 13 years and 1 month (8364 time stamps) as shown in Figure 4. The observation error standard deviation $\sigma^{\mathtt{R}} = 0.3\,\text{mm}$ was provided by the instrumentation engineers. Based

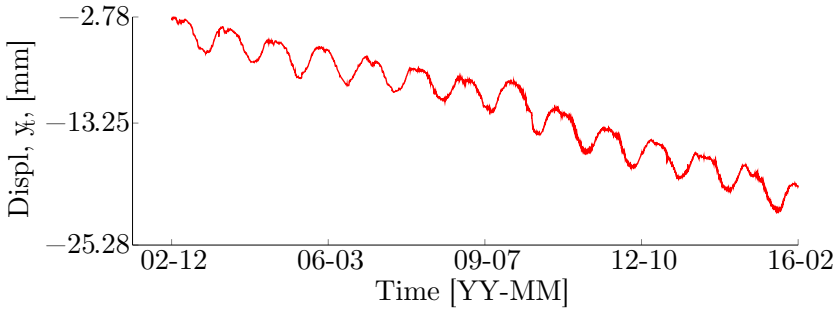

Figure 4: X-direction displacement data collected over the period of 13 years and 1 month.

on the raw data, one can observe a linear trend and a seasonal pattern with a period of one year. The seasonal pattern reaches its maximum during winter and minimum during summer and is *non-harmonic* because of the lack of symmetry with respect to the horizontal axis. The non-harmonic behaviour can be explained by the dependence of the displacement data on the water temperature [14, 34], where its variation is not harmonic due to the unbalanced duration between the reservoir warming and cooling periods. Figure 5 presents the time-step length for the entire dataset. Time-step length varies in the range between

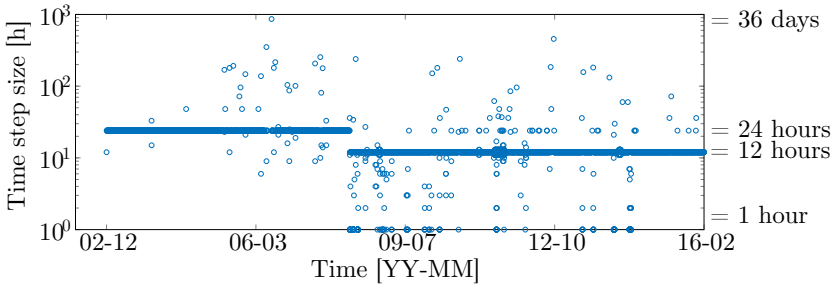

Figure 5: Time-step size is presented in a log scale.

1 hour to 36 days in which the two most frequent time steps are 12 and 24 hours. In order to adapt with the non-uniformity of time steps, the parameters need to be defined as a function of the time-step length where the *reference time-step* is selected by the most frequent one [25].

4.2 Model Construction

The probability of a state switch is estimated for two model classes representing respectively a state $s_t \in \{1 : Normal, 2 : Abnormal\}$. The set of components employed in each model

class is identical and follow

$$\mathbf{x}_t = \left[\underbrace{x_t^{\texttt{LL}}}_{\text{local level}}, \underbrace{x_t^{\texttt{LT}}}_{\text{local trend}}, \underbrace{x_t^{\texttt{LA}}}_{\text{local acceleration}}, \underbrace{x_t^{\texttt{T1,S1}}, x_t^{\texttt{T1,S2}}}_{\text{cycle, p = 365.24 days}}, \underbrace{x_t^{\texttt{T2,S1}}, x_t^{\texttt{T2,S2}}}_{\text{cycle, p = 182.62 days}}, \underbrace{x_t^{\texttt{AR}}}_{\text{AR}} \right]^{\texttt{T}} \tag{8}$$

In order to differentiate models, the *local acceleration* component for the Normal model class is forced to be equal to zero at every time step. This is done by assigning a value of zero to the line and row corresponding to the local acceleration component in the transition matrix $\mathbf{A}_t$ and model error covariance matrix $\mathbf{Q}_t$, see Appendix A for details. This constrain forces the Normal model class to have a constant velocity, i.e. acceleration = 0, while still allowing the Abnormal model to have a non-zero acceleration. The *baseline behaviour* is defined as the interaction of the local level, trend, and acceleration components. For both models, a superposition of two *hidden harmonic* components with a period of 365.24 and 182.62 days is used to describe the relationship between the displacement data and the hidden non-harmonic seasonal effect observed in the data, i.e. water temperature. Also, the AR component captures the time dependent model prediction errors.

The transition probability matrix is

$$\mathbf{Z} = \begin{bmatrix} Z^{11} & Z^{12} \\ Z^{21} & Z^{22} \end{bmatrix}, \tag{9}$$

where $Z^{ij} = \Pr(s_t = j | s_{t-1} = i)$ with $i, j = 1, 2$ is the prior probability of transitioning from a state i at time $t-1$ to a state j at time t. In order to be valid, this transition matrix must satisfy $\sum_j Z^{ij} = 1$. Given this constraint, only transition probabilities Z^{ii} need to be defined as unknown parameters to be learned using MLE.

As presented in Section 3, the matrix $\mathbf{Q}_t$ needs to be defined in the case where there is a state transition between the previous state i at time $t-1$ and the arrival state j at time t. For this case-study, the local acceleration in the normal model is forced to be equal to zero so that only the uncertainty on the local trend component in the baseline behaviour is considered. In the case where there is no state transition from the state i to the state j, the model classes depend only on the arrival states j at time t. Therefore, the matrices $\mathbf{Q}^{i(j),\text{baseline}}$ are defined as

$$\mathbf{Q}_t^{1(2),\,\text{baseline}} = \begin{bmatrix} (\sigma^{\texttt{LA}})^2 \cdot \frac{\Delta t^2}{20} & 0 & 0 \\ 0 & (\sigma^{\texttt{LTT}})^2 \cdot \frac{\Delta t^3}{3} & 0 \\ 0 & 0 & (\sigma^{\texttt{LA}})^2 \cdot \Delta t \end{bmatrix}$$

$$\mathbf{Q}_t^{2(2),\,\text{baseline}} = \mathbf{Q}_t^{2,\,\text{baseline}}$$

$$\mathbf{Q}_t^{2(1),\,\text{baseline}} = \begin{bmatrix} (\sigma^{\texttt{LT}})^2 \cdot \frac{\Delta t^3}{3} & 0 & 0 \\ 0 & (\sigma^{\texttt{LTT}})^2 \cdot \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Q}_t^{1(1),\,\text{baseline}} = \mathbf{Q}_t^{1,\,\text{baseline}}$$

where $\sigma^{\texttt{LA}} \in \mathbb{R}^+$ is *local acceleration standard deviation* for abnormal model, $\sigma^{\texttt{LT}} \in \mathbb{R}^+$ is *local trend standard deviation* for normal model, $\sigma^{\texttt{LTT}} \in \mathbb{R}^+$ is *local-trend transition* (LTT) *standard deviation* for the state transition models, and Δt is the time step at time t. The full matrix $\mathbf{Q}_t$ employed in this case-study is presented in Appendix A. In this case-study, we employ the UD method in the filter step presented in Equation 3.

4.3 Parameter Estimation

The convergence of the parameter optimization is reached when the log-likelihood between two consecutive loops satisfies

$$\|\text{log-likelihood}_n - \text{log-likelihood}_{n-1}\| \leq 10^{-7} \times \|\text{log-likelihood}_{n-1}\|,$$

where n corresponds to n^{th} optimization loop. The initial mean $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{\Sigma}_0$ for hidden states are estimated using the multi-pass presented in Section 3 using a period of 5 years (1694 data points) and the number of iterations $N = 30$. This period is selected because of the absence of the state switch which is causing numerical instabilities in the Switching Kalman Smoother estimations. The set of unknown parameters $\mathcal{P}$ is defined as follow

$$\mathcal{P} = \left\{ Z^{11}, Z^{22}, \phi^{\texttt{AR}}, \sigma^{\texttt{LT}}, \sigma^{\texttt{LA}}, \sigma^{\texttt{LTT}}, \sigma^{\texttt{T1}}, \sigma^{\texttt{T2}}, \sigma^{\texttt{AR}} \right\}, \tag{10}$$

where the possible range for each parameter is: *transition probabilities*; $Z^{ii} \in (0, 1)$, *autocorrelation coefficient*; $\phi^{\texttt{AR}} \in (0, 1)$, *local trend standard deviation*; $\sigma^{\texttt{LT}} \in \mathbb{R}^+$, *local acceleration standard deviation*; $\sigma^{\texttt{LA}} \in \mathbb{R}^+$, *transition local trend standard deviation*; $\sigma^{\texttt{LTT}} \in \mathbb{R}^+$, *harmonic-component standard deviations* for a period of 365.24 and 182.62 days; $\{\sigma^{\texttt{T1}}, \sigma^{\texttt{T2}}\} \in \mathbb{R}^+$ respectively, and *autocorrelation standard deviation*; $\sigma^{\texttt{AR}} \in \mathbb{R}^+$. Initial parameter values are estimated based on engineering heuristics so that

$$\mathcal{P}^0 \quad = \quad \left\{ 0.9999, 0.95, 0.986, 1.62 \times 10^{-8}, 4 \times 10^{-4}, 0.07, 5.1 \times 10^{-7}, 2.7 \times 10^{-7}, 0.03 \right\}.$$

This optimization procedure employs the entire dataset (8364 data points) for estimating the parameter values $\mathcal{P}^*$.

4.4 Results and Discussion

The set of parameters $\mathcal{P}$ is estimated with MLE for the entire dataset. The parameter calibration is done on a computer with 32 Gb of Random Access Memory (RAM) and Intel i7 processor. The computational time required for the calibration task is approximately an hour and a half. The optimal parameter values identified are

$$\mathcal{P}^* \quad = \quad \{1, 0.912, 0.996, 3.65 \times 10^{-6}, 2 \times 10^{-8}, 7.4 \times 10^{-4}, 0.062, 1.9 \times 10^{-6}, 2.78 \times 10^{-5}, 0.021\},$$

where the ordering of each parameter remains identical as in Equation 10. The log-likelihood associated with this set of parameters is $\ln p(\mathbf{y}_{1:T}|\mathcal{P}^*) = 1784.3$. Combining the BDLM framework with SKF serves two purposes: (1) it enables the detection of anomalies without triggering false alarms and (2) it decomposes the observations into their hidden components. Figure 6 presents the probabilities of each model class estimated at each time step. The method proposed identifies that there is an abnormal event occurring between July 8 and

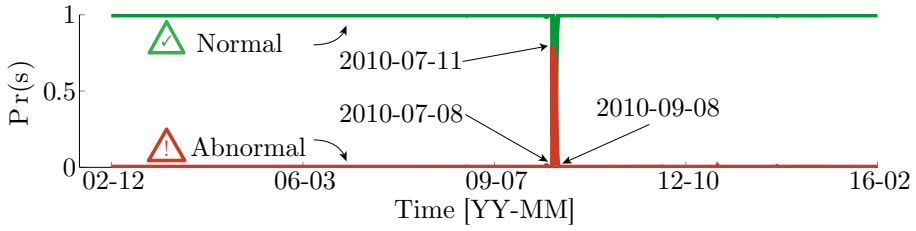

Figure 6: Probabilities of the two states are evaluated using SKF algorithm for the entire dataset.

11, 2010. This anomaly was caused by refection work that took place on the dam in early July. After the work was completed, the model identifies that the dam behaviour returns to a normal behaviour. This example of application demonstrates how anomalies can be detected without triggering any false alarm that would jeopardize the applicability of the approach.

Figure 7 presents the hidden components estimated for the entire dataset. The solid black line represents the mean values μ and its $\pm 1\sigma$ standard deviation interval is represented by the shaded region. Figure 7a, b, and c show a sudden change in the local level, trend and acceleration at the moment when the anomaly occurred. These three figures show how the baseline behaviour of the structure can be isolated from the effect of external factors. The external effect is modeled by a superposition of two harmonic components in Figure 7d and e. The local trend and local acceleration show a stable behaviour before the anomaly with estimated values of respectively $-2.0\,\text{mm/year}$ and $-0.0054\,\text{mm/year}^2$. During the abnormal event, local trend and acceleration components indicate a discontinuity before returning to a normal behaviour where the acceleration is zero. After the anomaly, it is estimated that the rate of change in the structure displacement has increased in magnitude from -2.0 to $-2.9\,\text{mm/year}$. Figure 7f shows that the autoregressive component, as expected, follows a stationary procedure. If it would not be the case, the non-stationary would indicate that either the component choice, or the optimal parameter identified are inadequate. Note that the sudden jumps in uncertainty bounds in Figure 7b and c are caused by prolonged duration without data as illustrated by the non-uniformity of time-steps in Figure 5.

This case-study illustrates the potential of a combination of the BDLM framework with SKF for detecting anomalies in the behaviour of structures. One limitation of the current method is that the unknown parameters associated with the models are assumed to be constant over time. Another limitation is that there is currently no quantitative guarantee that the approach is suited for other types of anomalies not similar to the case studied here. This aspect will need to be addressed in future work. The development of an *online anomaly detection* methodology is the subject of current research.

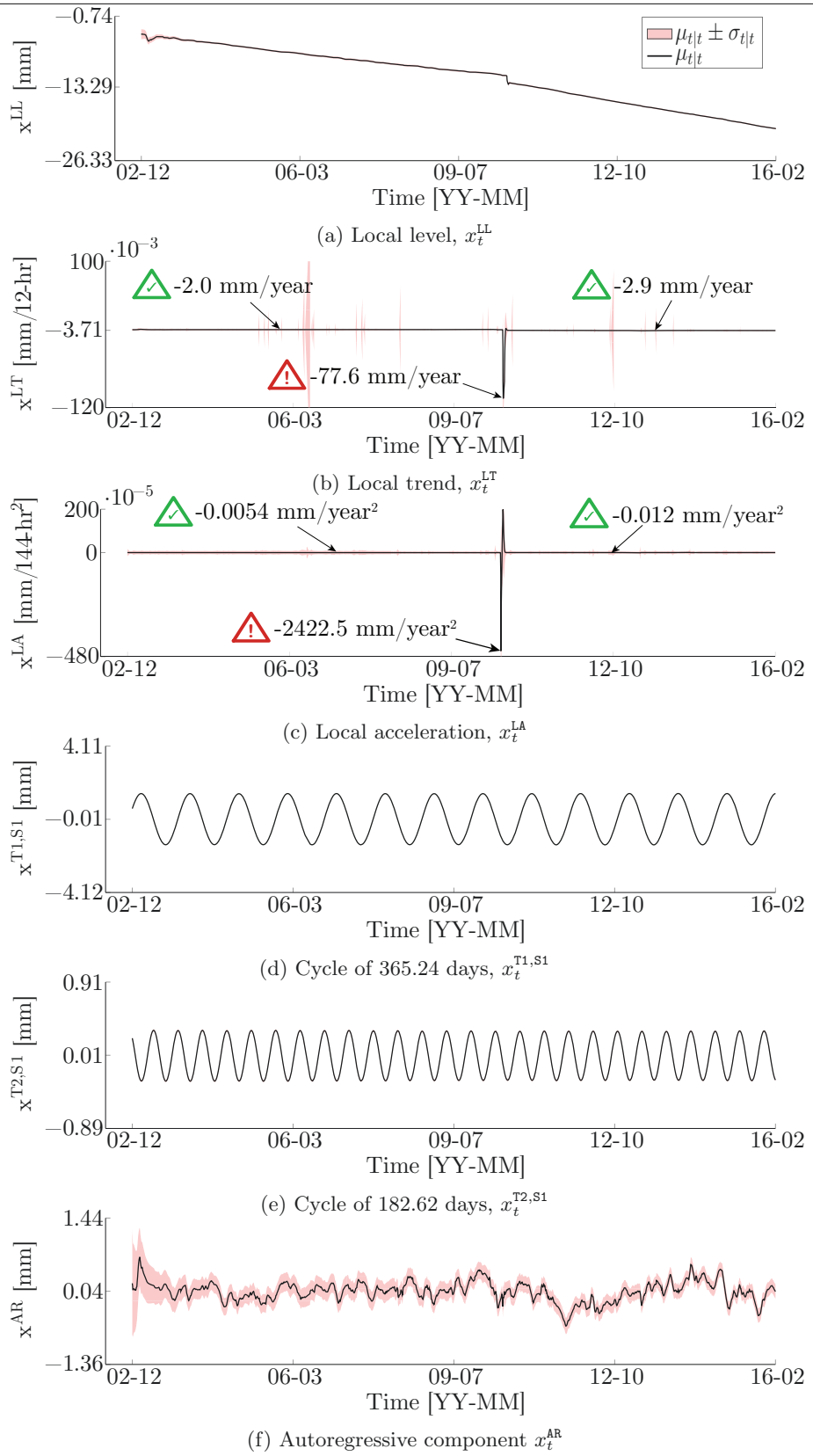

Figure 7: Expected values $\mu_{t|t}$ and uncertainty bound $\mu_{t|t} \pm \sigma_{t|t}$ for hidden components of a combination of models 1 and 2 are evaluated using SKF algorithm.

5 Conclusion

This paper presents a new approach combining the Bayesian Dynamic Linear Models framework with the Switching Kalman Filter theory for detecting anomalies of the behaviour structures. The key aspects are that (1) it enables early anomaly detection, (2) it is robust towards false alarms in real operation condition, and (3) it does not require labeled training data with normal and abnormal conditions. The approach is applied to the horizontal displacement data collected on a dam in Canada. In the case study considered, the method has shown that it was capable to detect the changes in the dam behaviour caused by the refection work. It also provided the specific information about the dam behaviour over time. This new approach offers a promising path toward the large-scale deployment of SHM system for monitoring behaviour of a population of structures.

Acknowledgements

This project is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2016-06405). The authors acknowledge the contribution of Hydro-Quebec who provided the dataset employed in this research. More specifically, the authors thank Benjamin Miquel and Patrice Côté from Hydro-Quebec for their help in the project.

References

[1] C. Macilwain, Out of service, Nature 462 (2009) 846.

[2] ASCE, 2013 report card for America's infrastructure, Tech. rep., American Society of Civil Engineers, Washington (2013).

[3] P. Cawley, R. Adams, The location of defects in structures from measurements of natural frequencies, The Journal of Strain Analysis for Engineering Design 14 (2) (1979) 49–57.

[4] R. Adams, P. Cawley, C. Pye, B. Stone, A vibration technique for non-destructively assessing the integrity of structures, Journal of Mechanical Engineering Science 20 (2) (1978) 93–100.

[5] F. Catbas, T. Kijewski-Correa, A. Aktan, Structural Identification of Constructed Facilities. Approaches, Methods and Technologies for Effective Practice of St-Id, American Society of Civil Engineers (ASCE), Reston, VA, 2013.

[6] J. Lynch, K. Loh, A summary review of wireless sensors and sensor networks for structural health monitoring, Shock and Vibration Digest 38 (2) (2006) 91–130.

[7] M. Pimentel, D. Clifton, L. Clifton, L.Tarassenko, A review of novelty detection, Signal Processing 99 (2014) 215 – 249. doi:10.1016/j.sigpro.2013.12.026.

[8] F. Salazar, R. Morán, M. Á. Toledo, E. Oñate, Data-based models for the prediction of dam behaviour: a review and some methodological considerations, Archives of Computational Methods in Engineering 24 (1) (2017) 1–21. doi:10.1007/s11831-015-9157-9.

[9] S. Ferry, G. Willm, Méthodes d'analyse et de surveillance des déplacements observés par le moyen de pendules dans les barrages, in: VIth International Congress on Large Dams, 1958, pp. 1179–1201.

[10] F. Lugiez, N. Beaujoint, X. Hardy, L'auscultation des barrages en exploitation au service de la production hydraulique d'électricité de france, des principes aux résultats, in: Xth International Congress on Large Dams, 1970, pp. 577–600.

[11] G. Willm, N. Beaujoint, Les méthodes de surveillance des barrages au service de la production hydraulique d'electricité de france, problèmes anciens et solutions nouvelles, in: IXth International Congress on Large Dams, 1967, pp. 529–550.

[12] P. Léger, M. Leclerc, Hydrostatic, temperature, time-displacement model for concrete dams, Journal of engineering mechanics 133 (3) (2007) 267–277. doi:10.1061/(ASCE)0733-9399(2007)133:3(267).

[13] P. Léger, S. Seydou, Seasonal thermal displacements of gravity dams located in northern regions, Journal of performance of constructed facilities 23 (3) (2009) 166–174. doi:10.1061/(ASCE)0887-3828(2009)23:3(166).

[14] M. Tatin, M. Briffaut, F. Dufour, A. Simon, J.-P. Fabre, Thermal displacements of concrete dams: Accounting for water temperature in statistical models, Engineering Structures 91 (2015) 26 – 39. doi:10.1016/j.engstruct.2015.01.047.

[15] J. Mata, Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models, Engineering Structures 33 (3) (2011) 903 – 910. doi:10.1016/j.engstruct.2010.12.011.

[16] H. Su, Z. Chen, Z. Wen, Performance improvement method of support vector machine-based model monitoring dam safety, Structural Control and Health Monitoring 23 (2) (2016) 252–266. doi:10.1002/stc.1767.

[17] L. Cheng, D. Zheng, Two online dam safety monitoring models based on the process of extracting environmental effect, Advances in Engineering Software 57 (2013) 48 – 56. doi:10.1016/j.advengsoft.2012.11.015.

[18] F. Salazar, M. Á. Toledo, J. M. González, E. Oñate, Early detection of anomalies in dam performance: A methodology based on boosted regression trees, Structural Control and Health Monitoring (2017) e2012–n/aE2012 stc.2012. doi:10.1002/stc.2012.

[19] V. Ranković, N. Grujović, D. Divac, N. Milivojević, A. Novaković, Modelling of dam behaviour based on neuro-fuzzy identification, Engineering Structures 35 (2012) 107 – 113. doi:10.1016/j.engstruct.2011.11.011.

[20] S. Gamse, M. Oberguggenberger, Assessment of long-term coordinate time series using hydrostatic-season-time model for rock-fill embankment dam, Structural Control and Health Monitoring 24 (1) (2017) e1859–n/a, e1859 STC-15-0220.R2. doi:10.1002/stc.1859.

[21] E. Carden, J. Brownjohn, Arma modelled time-series classification for structural health monitoring of civil infrastructure, Mechanical systems and signal processing 22 (2) (2008) 295–314.

[22] J.-B. Bodeux, J.-C. Golinval, Application of armav models to the identification and damage detection of mechanical and civil engineering structures, Smart materials and structures 10 (3) (2001) 479.

[23] J. Yang, S. Lin, H. Huang, L. Zhou, An adaptive extended kalman filter for structural damage identification, Structural Control and Health Monitoring 13 (4) (2006) 849–867. doi:10.1002/stc.84.

[24] J. Yang, S. Pan, H. Huang, An adaptive extended kalman filter for structural damage identifications ii: unknown inputs, Structural Control and Health Monitoring 14 (3) (2007) 497–521. doi:10.1002/stc.171.

[25] J.-A. Goulet, Bayesian dynamic linear models for structural health monitoring, Structural Control and Health Monitoring (2017) e2035–n/aE2035 stc.2035. doi:10.1002/stc.2035.

[26] K. Murphy, Switching kalman filters, Tech. rep., Citeseer (1998).

[27] W. Wu, M. Black, D. Mumford, Y. Gao, E. Bienenstock, J. Donoghue, Modeling and decoding motor cortical activity using a switching kalman filter, IEEE transactions on biomedical engineering 51 (6) (2004) 933–942.

[28] V. Manfredi, S. Mahadevan, J. Kurose, Switching kalman filters for prediction and tracking in an adaptive meteorological sensing network, in: Sensor and Ad Hoc Communications and Networks, 2005. IEEE SECON 2005. 2005 Second Annual IEEE Communications Society Conference on, IEEE, 2005, pp. 197–206.

[29] P. Lim, C. Goh, K. Tan, P. Dutta, Multimodal degradation prognostics based on switching kalman filter ensemble, IEEE transactions on neural networks and learning systems 28 (1) (2017) 136–148.

[30] K. Murphy, Machine learning: a probabilistic perspective, The MIT Press, 2012.

[31] D. Simon, Optimal state estimation: Kalman, H infinity, and nonlinear approaches, Wiley, 2006.

[32] S. Russell, P. Norvig, Artificial Intelligence, A modern approach, Prentice-Hall, 1995.

[33] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian data analysis, 3rd Edition, CRC Press, 2014.

[34] F. Salazar, M. Toledo, Discussion on "thermal displacements of concrete dams: Accounting for water temperature in statistical models", Engineering Structures (2015) –doi:10.1016/j.engstruct.2015.08.001.

Appendix A

The *transition matrix* ($\mathbf{A}_t$), the *observation matrix* ($\mathbf{C}_t$), the *observation error covariance matrix* ($\mathbf{R}_t$), and the *model error covariance matrix* ($\mathbf{Q}_t$) for normal model class and abnormal model class are defined following

Normal model class

$$\mathbf{A}_t^1 = \text{block diag}\left(\begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T1}} & \sin\omega^{\mathtt{T1}} \\ -\sin\omega^{\mathtt{T1}} & \cos\omega^{\mathtt{T1}} \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T2}} & \sin\omega^{\mathtt{T2}} \\ -\sin\omega^{\mathtt{T2}} & \cos\omega^{\mathtt{T2}} \end{bmatrix}, \phi^{\mathtt{AR}}\right)$$

$$\mathbf{C}_t^1 = [1,0,0,1,0,1,0,1]$$

$$\mathbf{R}_t^1 = \left[\left(\sigma^{\mathtt{R}}\right)^2\right]$$

$$\mathbf{Q}_t^{1(1)} = \text{block diag}\left(\left(\sigma^{\mathtt{LT}}\right)^2 \cdot \begin{bmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} & 0 \\ \frac{\Delta t^2}{2} & \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T1}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T1}}\right)^2 \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T2}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T2}}\right)^2 \end{bmatrix}, \left(\sigma^{\mathtt{AR}}\right)^2\right)$$

$$\mathbf{Q}_t^{2(1)} = \text{block diag}\left(\begin{bmatrix} \left(\sigma^{\mathtt{LT}}\right)^2 \cdot \frac{\Delta t^3}{3} & 0 & 0 \\ 0 & \left(\sigma^{\mathtt{LTT}}\right)^2 \cdot \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T1}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T1}}\right)^2 \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T2}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T2}}\right)^2 \end{bmatrix}, \left(\sigma^{\mathtt{AR}}\right)^2\right)$$

Abnormal model class

$$\mathbf{A}_t^2 = \text{block diag}\left(\begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T1}} & \sin\omega^{\mathtt{T1}} \\ -\sin\omega^{\mathtt{T1}} & \cos\omega^{\mathtt{T1}} \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T2}} & \sin\omega^{\mathtt{T2}} \\ -\sin\omega^{\mathtt{T2}} & \cos\omega^{\mathtt{T2}} \end{bmatrix}, \phi^{\mathtt{AR}}\right)$$

$$\mathbf{C}_t^2 = [1,0,0,1,0,1,0,1]$$

$$\mathbf{R}_t^2 = \left[\left(\sigma^{\mathtt{R}}\right)^2\right]$$

$$\mathbf{Q}_t^{1(2)} = \text{block diag}\left(\begin{bmatrix} \left(\sigma^{\mathtt{LA}}\right)^2 \cdot \frac{\Delta t^2}{20} & 0 & 0 \\ 0 & \left(\sigma^{\mathtt{LTT}}\right)^2 \cdot \frac{\Delta t^3}{3} & 0 \\ 0 & 0 & \left(\sigma^{\mathtt{LA}}\right)^2 \cdot \Delta t \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T1}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T1}}\right)^2 \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T2}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T2}}\right)^2 \end{bmatrix}, \left(\sigma^{\mathtt{AR}}\right)^2\right)$$

$$\mathbf{Q}_t^{2(2)} = \text{block diag}\left(\left(\sigma^{\mathtt{LA}}\right)^2 \cdot \begin{bmatrix} \frac{\Delta t^2}{20} & \frac{\Delta t^4}{8} & \frac{\Delta t^3}{6} \\ \frac{\Delta t^4}{8} & \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^3}{6} & \frac{\Delta t^2}{2} & \Delta t \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T1}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T1}}\right)^2 \end{bmatrix}, \begin{bmatrix} \left(\sigma^{\mathtt{T2}}\right)^2 & 0 \\ 0 & \left(\sigma^{\mathtt{T2}}\right)^2 \end{bmatrix}, \left(\sigma^{\mathtt{AR}}\right)^2\right)$$

where Δt is the time step at the time t.